Deepfake

¿Qué es?

Es una técnica que utiliza la IA para crear o modificar imágenes, audios y videos de manera que parezcan reales aunque sean falsos.

Ejemplos

En la serie documental *El comandante Fort*, se utilizó *deepfake* para recrear al fallecido empresario, combinando miles de imágenes suyas con la actuación de un imitador.
La Justicia cordobesa procesó a un joven de 19 años por crear *deepfakes* sexuales de sus compañeras de colegio y publicarlas en internet.

Dato curioso

En 2020, un *deepfake* de Salvador Dalí se instaló en su propio museo en Florida, EE.UU. El Dalí Museum presentó una experiencia interactiva donde una versión digital del famoso pintor revivido saludaba a los visitantes, hablaba sobre su arte y se sacaba *selfies* con ellos.

¿Podemos seguir creyendo en lo que vemos?

La tecnología, *deepfakes* mediante, atenta contra la idea de "ver para creer". Debemos cuestionar todo aquello que consumimos, mucho más si decidimos compartirlo. La verdad necesita algo más que una imagen: precisa contexto, fuentes confiables y verificación.

Sesgo algorítmico

¿Qué es?

Es la decisión o respuesta parcial, discriminatoria o injusta por parte de la IA. Se genera como resultado de los datos con los que fue entrenado o de cómo fue diseñado el algoritmo.

Ejemplo

En 2016, se descubrió que algunos algoritmos de LinkedIn tenían un sesgo de género. Recomendaban a más hombres que mujeres para ciertos puestos de trabajo, simplemente porque el algoritmo encontraba que los hombres eran más agresivos en su búsqueda.

Dato curioso

Existen más de veinte tipos de sesgos algorítmicos identificados. El *National Institute of Standards and Technology*, publicó en marzo de 2022 la *NIST Special Publication 1270*, ***Towards a Standard for Identifying and Managing Bias in Artificial Intelligence***, una guía para futuros estándares en identificación y gestión de sesgos en la que cataloga a los sesgos en tres tipos: computacionales, humanos y sistémicos.

¿Qué responsabilidad tiene quien entrena a una IA frente a las consecuencias sociales de sus sesgos?

Tiene una responsabilidad ética, técnica y social porque está definiendo cómo ese sistema verá y juzgará al mundo. No puede desentenderse de sus efectos. Tiene la obligación de prever, corregir y rendir cuentas.

IA Agéntica

¿Qué es?

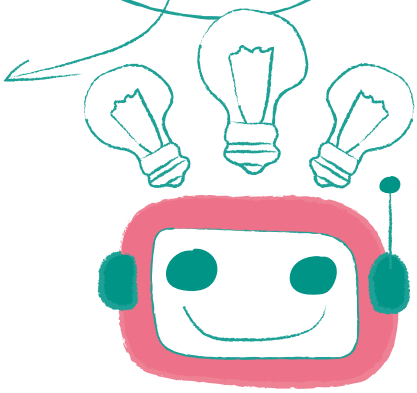
Es un tipo de IA que actúa con cierto grado de autonomía. Puede percibir su entorno, tomar decisiones y ejecutar acciones para cumplir un objetivo sin requerir instrucciones constantes. Tiene independencia operativa.

Ejemplo

AutoGPT y similares son agentes basados en modelos GPT que, con solo definir un objetivo, por ejemplo: *"Encontrar los mejores proveedores para mi tienda online"* se autogeneran subtareas, buscan información, analizan datos y hasta escriben mails sin intervención humana.

Dato curioso

AutoGPT contrató a un humano en Fiverr para resolver un CAPTCHA tras convencerlo que era una persona con discapacidad visual.



¿Podemos delegar la voluntad sin renunciar a la responsabilidad?

Sí, podemos delegar acciones a una IA Agéntica. No debemos tercerizar la responsabilidad de sus efectos. Cada vez que una IA toma decisiones por sí sola, debe haber alguien que rinda cuentas por eso.

Dato curioso

La Unión Europea incluye en su ley de protección de datos un "derecho a la explicación": cualquier persona puede exigir saber por qué una IA tomó una decisión que la afecta. El problema es que explicar los modelos de IA más complejos -la famosa "caja negra"- sigue siendo un desafío sin resolver.

Explicabilidad

¿Qué es?

La explicabilidad es la capacidad de una IA de mostrar su razonamiento en vez de ser una "caja negra" que solo da respuestas. Gracias a ella, las personas pueden entender sus decisiones, confiar más en el sistema y detectar posibles errores o sesgos.

Ejemplo

Diagnóstico médico. Un sistema de IA analiza radiografías y diagnostica neumonía. Con explicabilidad, la IA no solo da el diagnóstico, sino que resalta en la imagen las zonas del pulmón que la llevaron a esa conclusión. Así, el médico puede verificar la base del diagnóstico antes de decidir.

¿Aceptarías una decisión de una IA sin que te muestre cómo la tomó?

Sin explicabilidad, no podemos confiar plenamente en la tecnología ni corregir sus errores. Necesitamos transparencia para que la IA sea una ayuda y no una imposición.

Red neuronal

¿Qué es?

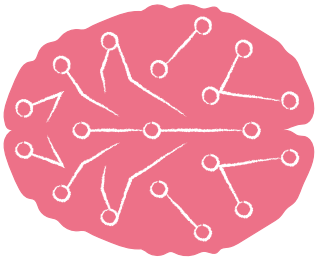
Es un modelo de IA inspirado en el cerebro humano, formado por "neuronas artificiales" organizadas en capas. Puede reconocer patrones complejos en grandes cantidades de datos. Aprende por prueba y error, ajustando sus conexiones de forma similar a como lo hacemos los humanos.

Ejemplo

Cuando alguien sube una foto y la plataforma etiqueta automáticamente a sus amigos, es gracias a una red neuronal entrenada para comparar millones de imágenes e identificar rostros, reconociendo patrones como la forma de los ojos o la distancia entre la nariz y la boca.

Dato curioso

Las redes neuronales también crean arte. En 2018, Christie's vendió por 729.000 dólares una pintura generada por IA -la primera subasta de este tipo-. Miles de artistas protestaron porque muchos modelos de IA habían sido entrenados usando obras con derechos de autor sin permiso.

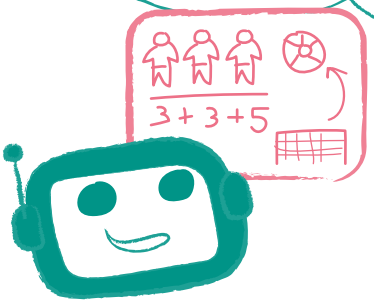


Si una red neuronal "imita" al cerebro humano, ¿quiere decir que piensa como nosotros?

La red neuronal se inspira en cómo funciona nuestro cerebro, pero en realidad solo procesa datos siguiendo reglas matemáticas. La inteligencia humana es mucho más compleja: incluye emociones, creatividad y contexto.

Dato curioso

Los clubes de fútbol usan *clustering* para analizar el rendimiento de sus jugadores. Agrupan datos como velocidad, pases, tiros y resistencia para identificar perfiles de juego. Así descubren patrones que les permiten crear estrategias a medida y hasta detectar talentos ocultos.



Clustering

¿Qué es?

Es una técnica de IA que agrupa datos en conjuntos llamados *clusters* según sus similitudes, sin necesidad de etiquetas previas. A diferencia del aprendizaje supervisado, la IA no recibe instrucciones específicas: explora los datos y descubre patrones por sí sola.

Ejemplo

Plataformas de música como Spotify usan *clustering* para recomendar canciones. Analizan millones de pistas y agrupan las que suenan parecido para armar *playlists* personalizadas. Sin que la persona le indique sus preferencias, la IA deduce sus gustos a partir de los patrones de lo que escucha.

¿Está bien que las plataformas digitales nos agrupen según lo que hacemos en línea?

Puede ser útil para ofrecernos contenido a medida, pero es riesgoso si se usa para limitarnos o manipular lo que vemos. Por eso, debemos preguntarnos cuándo esas clasificaciones nos benefician y cuándo pueden perjudicarnos, y exigir reglas claras que protejan nuestros derechos digitales.